Project Hydra:
Designing & Building a Reusable Framework for Multipurpose, Multifunction, Multi-institutional Repository-Powered Solutions

PAPER PROPOSAL for Open Repositories 2009

Repositories have proven themselves as powerful tools for managing digital content in many different contexts. But experience has also shown that there are real, practical limits in trying to extend a single repository solution to meet the manifold needs of most institutions for their full range of digital content and use cases. Relatively narrow and inflexible application front ends can be used to create single-purpose repository-powered solutions, but they do not lend themselves to being quickly and easily repurposed to meet variations in content type or user interactions.

There is a clear business need for a flexible, reusable application framework that can support the rapid development of multiple systems tailored to distinct needs, but powered by a common underlying repository. Recognizing this common need, Stanford University, the University of Hull and the University of Virginia are collaborating on "Project Hydra", a three-year effort to create an application and middleware framework that, in combination with an underlying Fedora repository, will create a reusable environment for running multifunction, multipurpose repository-powered solutions.

Multipurpose: Across the three institutions the partners are currently running more than a half dozen repositories, with known needs for more than a dozen more use cases and content types. The requirements and specification for Hydra functionality are based on a full range of use cases and content across all three institutions.

These include support for

- digital library use cases, including the accessioning, management and delivery of electronic content of a variety of formats, processed through complex workflows,
- institutional repository use cases, including deposit, management and dissemination of a of scholarly materials from scholars across the academic communities,
- personal repository use cases, providing a working space for scholars to help manage their individual or group research, through its complete lifecycle, and
- integration with other institutional systems, such as learning management systems and digital asset discovery and delivery applications.

To insure the Hydra initiative focuses on practical needs, we are developing the framework around specific real world applications. We will collaborate on generic "templates" for common solutions where a high degree of commonality exists across institutions, starting with a collaborative design yet individual deployments of electronic theses & dissertations (ETDs), a need largely similar across sites. In addition, to support a balance of reuse and local flexibility, each partner institution will deploy a locally distinct Hydra application served by their own Fedora instance. This is critical if Hydra is to support different classes of users across institutions, differences in workflows, or subtle variations in data models. In addition to ETD's, Hull will support a scholar's personal repository beneath its RepoMMan interface, UVa will manage and deliver digital manuscripts through a tailored application suite, and Stanford will manage the accessioning of PDF's and other simple digital objects through library workflows into its digital library.

Design Principles: The central design objective of the project is connoted by the name Hydra, with one body supporting many different heads. Using a common "body" comprising a Fedora repository and reusable application framework, institutions should be able to rapidly and easily deploy new applications with tailored interfaces.

To help achieve this high level of reuse and rapid time to deployment, the project team is following a number of complementary strategies by developing:

- reusable "lego bricks" of web services for content, metadata and repository management,
- flexible content models that rely on a combination of simple content objects (e.g., for images) and aggregation objects that link simple objects together into units of meaning,
- a series of simple disseminators that return the contents of datastreams, and that can be widely used to leverage the Hydra content models,
- an easily customized visual design, to support localization and branding through simple means like configuration and alternate stylesheets, and
- a toolkit of reusable user interaction widgets (such as upload, search, or editing panels) that can be easily assembled to meet new needs, or adopted "out of the box" in user-tested templates for common functions.

In aggregate, these components will help provide an "out-of-the-box" solution for those that need a repository up and running quickly, with minimal customization. It will also provide a rich set of tools and components that will enable institutions with needs for tailored or highly customized solutions to assemble one relatively easily, compared to building it from scratch.

Technology Stack: To support this rapid application development environment, the team has settled on a technology stack that comprises:

- Fedora 3.x
- ActiveFedora, a Ruby gem for creating and managing objects in Fedora, (developed by MediaShelf, LLC),
- the Hydra client, a Ruby on Rails application that serves as the core of the front end of the solution, integrated with
- Blacklight, a faceted browser based on Ruby on Rails and Solr (developed by University of Virginia),

The choice of Ruby on Rails as a development platform for Hydra was not a difficult one for the team, quite apart from facilitating the choice of Blacklight for discovery. Ruby on Rails' emphasis on convention over configuration provides its adopters with both a significantly reduced time to deploy new functionality, and a streamlined ongoing maintenance environment, for reducing the total cost of ownership over the long term. And so far it has lived up to its adherent's claims of promoting "programmer happiness".

Multifunction: Based on the experience of the Universities of Hull and Virginia, the team has identified five core functions that the initial Hydra client will support. These will provide feature parity with Muradora (developed at the Macquarie University in Sydney), which along with the RepoMMan toolset has provided the technology base for Hull's current repository system. These functions are:

- *Deposit*, – uploading simple or multi-part content
- *Manage* – including editing metadata, and setting permissions
- *Search* – based in SOLR using Blacklight UI to support user discovery
- *Browse* – by collections as well as other arbitrary or ad hoc attributes
- *Deliver* – via a combination of embedded Hydra functionality, Fedora disseminators, and third party applications

Three additional middleware capabilities will round out the longterm functionality of the Hydra solution. These are:

- Hooks to institutional *LDAP* directories, to provide *authentication services*
- Integration with a robust *permissions / authorization* service that interoperates with the underlying Fedora repository (i.e., not just at the Hydra application layer)
- Local deployments of *workflow* engines or strategies to support multistep processes comprising both human and automated steps in variable sequences.

Multi-institutional: The Hydra project is a collaboration of the three principal institutional partners, plus Fedora Commons, and also actively leveraging Media Shelf technologies. In addition, the collaboration is working with a small group of advisory institutions to ensure that its work is potentially of wide application. The effort is in its first of three years, with the primary goal of Year 1 being to establish a reusable application front end for a handful of content types and use cases that work across all three partner sites. Years 2 and 3 will tackle a greater diversity of content types, plus full integration with a robust security & permissions engine (still to be determined), plus the packaging of templates and application elements into "solutions bundles" for the wider Fedora community.

This presentation will provide demonstrations of the work done to date, including of the prototype ETD application, as well as the set of content models and disseminators that the project has defined so far. The presentation will also present links to the project's publicly accessible documentation and open source code, as well as solicit the constructive input from community members who may be interested in the project or its outcomes.